

Improving the quality of user generated data sets for activity recognition

Nugent, Chris D.; Synnott, Jonathan; Gabrielli, Celeste; Zhang, Shuai; Espinilla, Macarena; Calzada, Alberto; Lundström, Jens; Cleland, Ian; Synnes, Kåre; Hallberg, Josef; Spinsante, Susanna; Ortiz Barrios, Miguel Angel

Abstract

It is fully appreciated that progress in the development of data driven approaches to activity recognition are being hampered due to the lack of large scale, high quality, annotated data sets. In an effort to address this the Open Data Initiative (ODI) was conceived as a potential solution for the creation of shared resources for the collection and sharing of open data sets. As part of this process, an analysis was undertaken of datasets collected using a smart environment simulation tool. A noticeable difference was found in the first 1–2 cycles of users generating data. Further analysis demonstrated the effects that this had on the development of activity recognition models with a decrease of performance for both support vector machine and decision tree based classifiers. The outcome of the study has led to the production of a strategy to ensure an initial training phase is considered prior to full scale collection of the data.

Keywords

Activity recognition, Data driven classification, Data validation, Open data sets